

Exploration des pratiques de validation des instruments de mesure : une revue systématique de la littérature

Iris Bourgault Bouthillier, Michel Rousseau et Tristan Milot
Université du Québec à Trois-Rivières



Contexte

Plusieurs études ont été menées dans le but de décrire les pratiques de validation des instruments de mesure (Hogan et Agnello, 2004; Cizek et al., 2010; Shear et Zumbo, 2014; Hubley, et al., 2014). Les études publiées sur le sujet sont souvent axées sur des revues scientifiques en particulier, ou encore sur des domaines ou sur des instruments très précis (Lyons-Thomas et al., 2014), ce qui limite la possibilité d'avoir un portrait plus global des méthodes de validation utilisées.

L'objectif de la présente étude est de décrire les pratiques des chercheurs lors d'études de validation des instruments de mesure en documentant 1) les différents types de preuves de fidélité et de validité rapportés et 2) les analyses statistiques utilisées pour récolter les preuves. Pour ce faire, nous avons analysé 550 articles scientifiques portant sur la validation d'instruments de mesure et publiés entre 2002 et 2012.

Méthode

50 articles par année de publication (de 2002 à 2012), pour un total de 550 articles ont été sélectionnés aléatoirement dans PsychINFO à l'aide des mots clés *test validity*

Les articles non disponibles en version électronique ont été exclus de l'échantillon et remplacés (n=135)

Vérification de l'admissibilité des articles en texte intégral
Critères d'admissibilité: 1) article de validation d'un instrument de mesure, 2) instrument portant sur un concept psychologique, 3) en anglais ou en français

Les articles inadmissibles ont été exclus et remplacés. (n=167) L'admissibilité des nouveaux articles a aussi été vérifiée.

Article inclus dans les analyses quantitatives et qualitatives (n=550)

Les 550 articles retenus ont été lus et codifiés par trois assistantes de recherche dans une grille de codification recensant les différents types de preuves de fidélité et de validité qui y étaient utilisés.

Résultats

Proportion d'utilisation de chaque type de preuve de fidélité

	n	%
1. Cohérence interne	392	71
2. Stabilité temporelle	150	27
3. Fidélité interjuges	56	10
4. Équivalence de contenu	5	1

Proportion d'utilisation de chaque type de preuve de validité

	n	%
1. Relations à d'autres variables :		
1.1 corrélationnelles	354	64
1.2 comparaison entre des groupes	116	21
1.3 critériées	94	17
1.4 études expérimentales	20	4
2. Structure interne	312	57
3. Contenu	37	7
4. Processus de réponse aux items	14	3
5. Conséquences du testing	0	0

Analyses statistiques - Points saillants

Cohérence interne 95% des preuves sont calculées à l'aide de l'alpha de Cronbach	Structure interne 29% d'analyses factorielles exploratoires 14% d'analyses factorielles confirmatoires 13% présentent les deux types d'analyses
--	---

Discussion

Fidélité:

- La stabilité temporelle est moins utilisée, pourtant, pratiquement tous les instruments de mesure sont sujets aux erreurs de mesure aléatoires dues aux circonstances personnelles et aux conditions d'administration dont l'étendue peut être estimée grâce à la stabilité temporelle.
- Ne demandant qu'une seule passation d'un test, la cohérence interne est beaucoup plus facile à mesurer que la stabilité temporelle qui demande deux passations. Cela pourrait expliquer que ce deuxième type de preuve de fidélité soit moins utilisé.
- L'alpha de Cronbach est une méthode largement répandue pour calculer la cohérence interne. D'autres méthodes plus optimales existent, comme le Lambda2 de Guttman par exemple (Bertrand & Blais, 2004).

Validité:

- Les preuves les plus utilisées sont calculées à l'aide de méthodes plus objectives (indices statistiques), alors que les preuves moins utilisées requièrent un jugement subjectif de la part d'experts (preuves basées sur le contenu) ou de participants (preuves basées sur le processus de réponse aux items).
- Les études expérimentales sont probablement moins présentes car elles sont souvent utilisées dans des contextes bien précis, comme lorsqu'on veut vérifier la capacité d'un test à mesurer les fluctuations d'un trait.
- Nous n'avons repéré aucune preuve basée sur les conséquences du testing. Ce constat est cohérent avec les études de Hogan et Agnello (2004) et de Cizek et al. (2010) qui n'en ont pas trouvé non plus.
- Les analyses factorielles exploratoires sont les plus utilisées pour déterminer la structure interne et ne devraient normalement pas constituer à elles seules une preuve de validité.

Conclusion

En somme, les études de validation d'instruments de mesure ne rapportent pas beaucoup de preuves de fidélité et de validité. Certains types de preuves semblent prédominer dans les pratiques de validation, au détriment de l'utilisation d'autres preuves, souvent tout aussi pertinentes à démontrer.